HL7 Clinical Genomics Weekly Call - May 31, 2016

Attendees

1. Bob Milius - NMDP - bmilius@nmdp.org
2. Bret Heale - Intermountain Healthcare - bheale@gmail.com
3. Jonathan Holt - SeqTechDx - jholt@seqtechdx.com
4. Perry Mar - Partners HealthCare System - pmar@partners.org
5. Gaston Fiore - BCH - gaston.fiore@gmail.com
6. Kevin Power - Cerner - kpower@cerner.com
7. Joseph Kane - Epic - jkane@epic.com
8. Joel Schneider - NMDP - jschneid@nmdp.org
9. Bob Freimuth - Mayo Clinic - freimuth dot robert at mayo dot edu
10. JD Nolen - Cerner - johndavid.nolen@cerner.com
11. Grant Wood - Intermountain
12. Andrea Pitkus, IMO, apitkus@imo-online.com
13. Brett Johnson - icanbrj@gmail.com
14. David Kreda - david.kreda@gmail.com (joining late!)
15. Scott Bolte - Scott.Bolte@gmail.com

Discussion

- Minutes approval
    - May WGM - Montreal
        - http://wiki.hl7.org/index.php?title=File:HL7CG_WGM_May2016_Montreal_Minutes.pdf
        - Motion to accept - Gil
        - 2nd - Bob F
        - Discussion -
        - yea/nay/abstain = 8 / 0 / Scott B, Gaston, Bret H. , Perry Mar,
        - results - minutes accepted
- Brief updates
    - ClinGen/ClinVar -
        - Continues to plug away at assertion model
        - There will be a meeting later this week with reps from ClinGen, GA4GH, HL7, NCBI in Boston to discuss how we can harmonize models and terminologies related to genetic variants
        - Goals to develop a shared model and technical spec, common terminology
        - Reps from HL7 world will be there
    - GA4GH -
        - Gil - see above
        - Started with fhir global alliance synergy project, looked at variant level. Other orgs were added, focus recentered around variants, June 2-3,
        - Bob M - will results be released (minutes/artifacts)?
        - Bob F - one of goals will be sharable model,
        - Bob F - BRIDG could use this model, but question of governance is still to determined
        - Gil - this overlaps with Information Model meeting, will there be an IM meeting this week?

        - Bob F - figuring it out will let everyone know
    - National Academies-
        - JD - DIGITizE: Looking for a long-term home and exploring possible grant funding to secure the long-term future of the group.
        - Grant - hoping to have pilot moving forward in June
    - other -
        - none

- Definitions - see chat.fhir.org, genomics stream
    - Sequenceontology.org for many, but not all terms
    -
- FHIR Deadlines
  For our Work Group to have anything in the STU3 ballot, we have several activities we must promptly address due to deadlines set by the HL7 FHIR Management Group. The key deadline is Wednesday, June 1, the day after our next CGWG call.

  By June 1, in particular, we must have voted for certain proposals that must be document in an officially sanctioned format. These proposals are forward-looking assertion that we are going to work to have certain actual specifications ready in the relative near future, some number of weeks before the September WGM. In particular, here are the individual proposals that all work groups, including ours, would need to get done (if not already done and approved at an earlier date) by June 1 to be considered eligible for the STU3 ballot process:

  1. FHIR Resource Proposals

  2. FHIR profile Proposals

  3. FHIR Implementation Guide Proposal

  Finally, all work groups, ours included, need to indicate no later than June 1 if they will be participating in the September FHIR Connectathon.

  Re (1). We approved the Sequence Resource proposal in 2015. Again, not the actual specification, which we continue to work on, but merely the go-ahead for that resource. We will eventually vote again on the exact specification that would part of STU3.

  Re (2). We must promptly consider proposals for each of the profiles. For that, we have drawn up a proposal for each, see:

  http://wiki.hl7.org/index.php?title=ObservationGenetics_FHIR_Profile_Proposal

  http://wiki.hl7.org/index.php?title=DiagnosticReportGenetics_FHIR_Profile_Proposal

http://wiki.hl7.org/index.php?title=DiagnosticOrderGenetics_FHIR_Profile_Proposal

http://wiki.hl7.org/index.php?title=FamilyMemberHistoryGenetics_FHIR_Profile_Proposal

http://wiki.hl7.org/index.php?title=SequenceConsensusSequenceBlock_FHIR_Profile_Proposal

http://wiki.hl7.org/index.php?title=DiagnosticReportHlaResults_FHIR_Profile_Proposal

Examples of approved profile proposals are below (link given by Lloyd):
http://wiki.hl7.org/index.php?title=Category:Approved_FHIR_Profile_Proposal

We hope that a majority will approve the go-ahead for proposals for these profiles. The current profiles work-in-progress for which the proposal apply are, of course, presently set out at the staging site (FHIRgenomics.org). For voting, we propose that each proposal be a separate google poll that closes at the beginning of our CGWG call on Tuesday, May 31.

These profile proposals correspond to proposals that we were polling last week for a FHIR Maturity Model target at the end of December 2016. That said, based on polling last week, the FMM target was 3 except for the HLA/consensus block and Family Hx, each of which tied between 1 and 3. That vote can also stay open through the beginning of our CGWG call on Tuesday, May 31.
See:
https://docs.google.com/forms/d/1nCWreYHSDAr-YH7wnDJdNb8S2uuyYkbmRRQUBNbZFos/viewanalytics

Re (3). We must promptly consider a proposal to produce an Implementation Guide (IG). Again, the proposal is here (http://wiki.hl7.org/index.php?title=Genomics_FHIR_Profile_Proposal) and its table of contents recaps the list of resources and profiles, which can be amended depending upon the vote in (2) and in later final commit votes, etc.

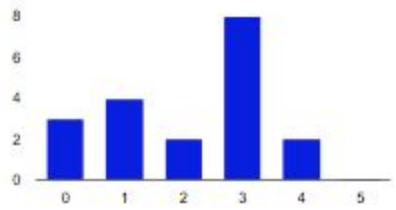
FMM Poll Results - 19 respondents at 9:30am
Simple majority (50+%) is needed to pass = 10/19
To say we've achieved consensus requires 66+% of the vote = 13/18

Resource/Profile	***FMM with more than 50%***	***FMM with at least 66% (consensus)***	***FMM with greatest number of votes***

	(majority)		
Observation for Genetics	3	1	3
Diagnostic Order for Genetics	3	1	3
Diagnostic Report for Genetics	3	1	3
Family Member History for Genetics Analysis	2	1	1, 2, 3 (tie)
Consensus Sequence Block	1	1	1
HLA Genotyping Results	1	1	1
Sequence	2	2	3

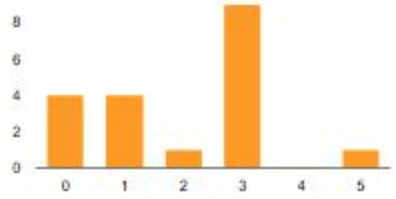
Profiles

Observation for Genetics



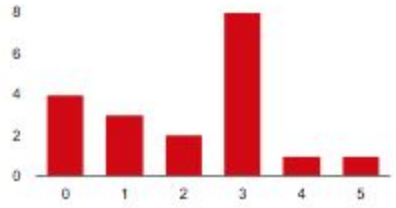
0	**3**	15.8%
1	**4**	21.1%
2	**2**	10.5%
3	**8**	42.1%
4	**2**	10.5%
5	**0**	0%

Diagnostic Order for Genetics



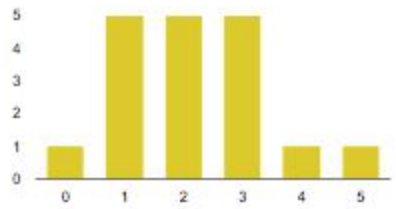
0	**4**	21.1%
1	**4**	21.1%
2	**1**	5.3%
3	**9**	47.4%
4	**0**	0%
5	**1**	5.3%

Diagnostic Report for Genetics



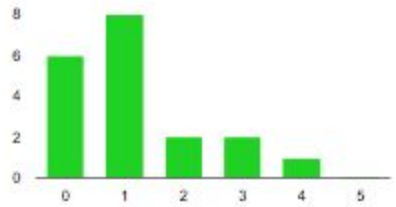
0	**4**	21.1%
1	**3**	15.8%
2	**2**	10.5%
3	**8**	42.1%
4	**1**	5.3%
5	**1**	5.3%

Family member history for genetics analysis



0	**1**	5.6%
1	**5**	27.8%
2	**5**	27.8%
3	**5**	27.8%
4	**1**	5.6%
5	**1**	5.6%

Consensus Sequence Block



0	**6**	31.6%
1	**8**	42.1%
2	**2**	10.5%
3	**2**	10.5%
4	**1**	5.3%
5	**0**	0%

HLA Genotyping Results

Motion to accept majority level (50+%) - First results column of table above - Gil
Second - Grant
Discussion - clarification of FMM levels
Vote = unanimous / 0 / 0

IG and Resource/Profile proposals

Profiles

Observation for Genetics

Yes	9	100%
No	0	0%

Diagnostic Order for Genetics

Yes	9	100%
No	0	0%

Diagnostic Report for Genetics

Yes	9	100%
No	0	0%

Family member history for genetics analysis

Yes	9	100%
No	0	0%

Consensus Sequence Block

Yes	8	100%
No	0	0%

HLA Genotyping Results

Yes	8	100%
No	0	0%

Results = accept doodle results
All will be submitted

Other Discussions

- Phase
- Jon - FastQ has additional characters ambiguous characters, for instance R represents an A or G at the locus, Unfortunately this handles ambiguous calls not really meant for ambiguous phase.
- Three levels of phase resolution
    - Technical - long range / single molecule sequencing, still an observation, but one that looks into the raw data (fastq) files for co-occurring bases on a read segment (typically around 200bp at most, but some technologies are many more hundreds long)
    - Imputation based on frequencies of co-occurring sequences
    - Family history = deduction (will need to point to genetics family history sequences)
- Gil - sequences can be put into sets

Gaston's on-the-fly notes are below. Please add and fix accordingly!

Two things: are either on phase or not on phase. Should we have a flag to indicate these two sequences or two alleles are in phase or not?

FastQ format, additional N for no calls, D stands for A or C, either ambiguity or indicate phasing with the characters. Bob says this probably works in the Sequence. I'm thinking more Sequence Exon 1, and Exon 2, haven't sequenced the introns, what to say whether they're in phase or not. More at the sequence level, not at the nucleotide level. All of the variants within that Exon are in phase with each other and also in phase with the variant in the other exon down? Two levels of phasing. Phasing issue within a sequence block isn't as acute as it used to be. But if I'm targeting 2 exons, typical in HLA, and I find 4 exons, i get two from one and 2 from the other. Doing some extra sequencing to extend into the introns. But i also want to know if there's no phase.

What are the 3 options? You cannot just say these are in phase. You also have to say why they're in phase. This corresponds to the observation. Now, what's required in the raw data to support that observation? In HML we have a thing called phase_set. If I have 4 exons, don't know phase between them, have identifiers for these. But if I know 2 of them are in phase, I can say only 2 are in phase and may imply that if 2 are in same phase, the other 2 are also probably in same phase.

Bob F: want to be able to create individual statements of phase between any 2 or more variants present or sequences . 2 or more observed alleles at a known variant site. it may not be a single variant site. 2 variant sites on the same gene. its more than just looking at a chunk of sequence and comparing at another chunk of sequence. it's to be able to a certain number of variant alleles are known to be in phasing on the same strand of dan and other variant alleles have undefined phase. starting to sound an interpretation or observation that is made multiple times a tteh level of individual variant alleles.

3 cases: Know the sequence, Look at imputation, or family history.

It doesn't have to be on the same strand either. The observation could draw on raw data from sequence. If it's one large sequence, an entire chromes, then this on sequence resource. but if we're doing fragmented sequencing, today's technology, interpretation based on observation.

Single nucleotide ambiguity and recording that we can start here. But if we want to compare different regions. That is the standard with fastq format.

If we don't know a phase between exons, and there are alleles, put allele assignment as a list of possibilities, it could be this genotype, or this genotype, etc. that's how we would report that. but for the sake of saying it could be this, we think it's this because we're assuming this phase, or it's this …

Two variants. Go back to fastq file. These two are in phase because two reads have these together. Easy call. How do we represent how this call is being made? Have 2 exons that are in phase, sequence across, that's an observation that goes back to fast file, raw, raw data, to make the observation. Distinction between statement of phase and method you used to arrive at that. They're intrinsically tied together. Different levels of belief why they're in phase.

Is it an observation attribute around 2 or more sequences? Have to be able to group 2 or more structures and put an attribute around those sequences. Variant 1..* within the sequence resource. See examples of instances on these things.

There's no reason why cannot have an observation for the sequence being called, and another one for the phase? Can I have in parallel? 3 sequence observation attached to 1 phase observation. DR is a terminal state. You cannot nest those. You cannot tie an allele call to a DR. In the clinFHIR tutorial you can make your own profile. What are the technical limitations of this? You can possibly create an observation with multiple observations that have what you're looking for.

Summary: it's doable. have to work the logistics of different observations. one observation related to variant. another observation related to the phasing with each referencing Sequence Resource(s)